

Three Bead Rotating Chain model shows universality in the stretching of proteins

Hamed Seyed-allaei

*International School for Advanced Studies (SISSA),
via Beirut 4, 34014 Trieste , Italy and
INFM, Trieste, Italy**

Abstract

We introduce a model of proteins in which all of the key atoms in the protein backbone are accounted for, thus extending the Freely Rotating Chain model. We use average bond lengths and average angles from the Protein Databank as input parameters, leaving the number of residues as a single variable. The model is used to study the stretching of proteins in the entropic regime. The results of our Monte Carlo simulations are found to agree well with experimental data, suggesting that the force extension plot is universal and does not depend on the side chains or primary structure of proteins.

PACS numbers: 87.14.Ee, 87.15.Aa, 87.15.La, 82.37.Rs

Keywords: Modeling and Simulation of Proteins, Single Molecule Experiments

*Electronic address: allaei@sissa.it

I. INTRODUCTION

Accurate modeling of protein structure and dynamics is an immense challenge in the biological sciences, and has elicited intense interest among physicists. Many models have been developed, from first principle studies to coarsegrained phenomenological theories. A major issue is the tradeoff between inclusion of microscopic details, and computational tractability and efficiency. It is hoped that the latter can be improved by removing select microscopic details which do not affect the accuracy of the results.

In this work we introduce a model of the protein backbone and apply it to the problem of protein stretching, for which many experimental results are available. This allow us to model the backbone of proteins accurately without paying attention to effective interactions or force-fields.

A. Experiments

We know the elastic property of single molecules by atomic force microscopy (AFM) and optical tweezers [1, 2, 3, 4, 5]. One of the most studied proteins by these experimental methods is titin, which plays an important role in the elasticity of muscles [6]. The interesting segment of titin is the I-band, which is made of similar modules (also called repeats or domains) with an immunoglobulin-like (Ig) structure [6, 7]. A single folded module can resist against pulling below a threshold. If the force exceeds the threshold, the domain will unfold. The unfolded domain will refold into its native state if the force is removed [1].

If we pull a chain of folded Ig domains slowly, the necessary force to keep the protein extended will increase until one of the modules unfolds. This is called an unfolding event, and after that we can keep the protein extended with a much weaker force. As we increase the extension, the modules unfold one by one until we get a fully extended chain. If we plot force versus extention during the stretching, we will see a saw-tooth pattern (Fig. 5) [1]. The stretching of titin also has been studied by different theoretical methods like molecular dynamic simulation [8, 9], lattice models [10] and the thick chain (TC) model [11].

B. Models

Many theoretical models have been developed to study macromolecules and proteins. Among them, Freely Jointed Chain (FJC), Freely Rotating Chain (FRC) and Worm Like Chain (WLC) are the most famous[12, 13, 14]. They are used intensively to model the backbone of proteins by considering the protein as a chain of only C_α atoms, and usually using an effective interaction among C_α atoms to recover the neglected details. We argue here that these models are not accurate enough to model protein backbone, and they usually have at least one free parameter in addition to the chain length N (for example, the persistence length) that should be calculated by fitting of theories to experiments. Instead of using such a parameter, we can simply use the well known geometric details of peptides.

Worm Like Chain is used to study the elasticity of macromolecules [14]. It is simple and works fine whenever we can apply continuum approximation. It is good especially for very large and stiff macromolecules like DNA [15, 16], whereas it becomes physically unjustifiable when applied to more coarse-grained and more flexible chains like proteins. For instance, in stretching of titin's Ig domain, the persistence length of the related WLC is ~ 0.4 . Although this value is about the size of a single peptide unit[1], WLC is frequently used in literature to fit and explain experimental data because a better alternative had been absent.

Freely Jointed Chain is a chain of monomers with fixed bond length [12]. In this model, bonds' angles can have any arbitrary value, which is not the case in real proteins [17]. However, Freely Jointed Chain has an advantage over more detailed models as it has an exact solution in the case of stretching.

Freely Rotating Chain has one more constraint than the FJC. The angles are rigid and do not change [13]. Freely Rotating Chain behaves more like real proteins compared to FJC. This model has been used to study different aspect of proteins and polymers including our problem of interest [18]. However, it does not cover all the features of bond angles because the angle between $C_\alpha - C_\alpha$ bonds is not completely rigid in real proteins (table II) [17].

In our model, we consider carbon and nitrogen atoms on the backbone (C_α, C and N) and we follow the Ramachandran picture. This can be considered as a fine-grained version of a FRC or as a coarse-grained version of a Four Bead model. Our model is the simplest model that includes all the geometric properties of the backbone of proteins, and it does not need any fitting parameters.

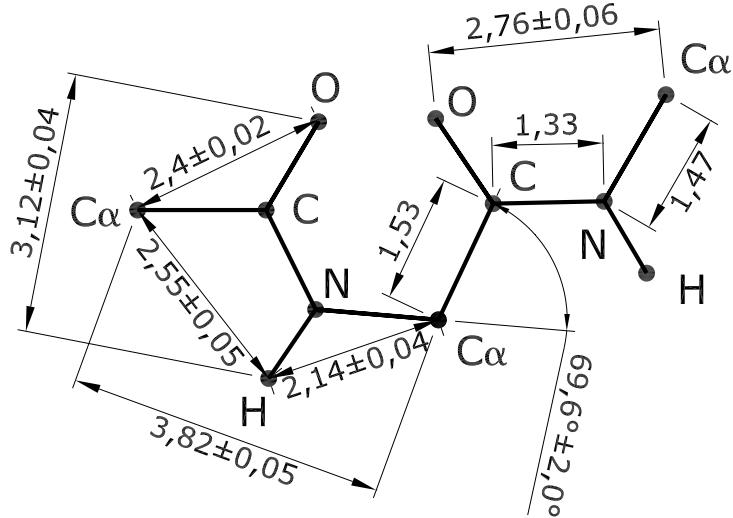


FIG. 1: Peptide units and their average dimensions (nanometer) and average angles, from Protein Data Bank (PDB).

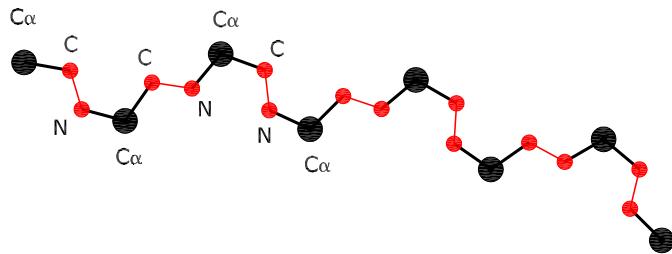


FIG. 2: (Color on-line) A Three Bead Rotating Chain. The rotations are done around $C_\alpha - C$ and $C_\alpha - N$ (bold black lines). We do not permit any rotation around $C - N$ bonds (red lines in color prints or thin gray lines that connect smaller circles in black and white prints).

In this work we are interested only in the geometric details rather than the effective interactions so we can keep the model simple and clear; therefore, we only study the mechanical properties of proteins in the entropic regime and compare the results with experiments.

II. MODEL AND SIMULATION

We use Freely Rotating Chain model with a small variation: some bonds can not be a pivot. If we show the chain as a binary sequence of 0 and 1 in which 1 represents a bond

that can be a pivot for rotations and 0 as a bond that can not be a base for rotations, then our chain will be presented as:

$$[101]_n \quad (1)$$

where $[\dots]_n$ means n repeats of the argument. This is equivalent to considering C and N atoms in addition to C_α atoms and using a Ramachandran picture so that the chain only rotates around $C_\alpha - C$ and $C_\alpha - N$ bonds [17] (Fig. 2). Since the chain does not rotate around $C - N$ bonds, we know the structure by only a pair of angles (ϕ, ψ) for each C_α junction. These angles are known as Ramachandran angles [17].

This is the simplest model that covers Ramachandran angles which has more details than FRC, and it is simpler than the Four Bead model[19, 20, 21].

To set distances and angles in our simulation, we analyzed the Protein Data Bank (PDB) and calculated the averages of desired quantities over all polypeptide chains in the database (72212 peptide units) . The results are shown in Fig. 1 and more details are in tables I and II.

The largest rigid distance over the backbone of proteins is the one between two sequential C_α atoms; therefore, a good approximation is to consider a chain of C_α atoms with fixed bond length. On the other hand, it is not as good to assume proteins act under the Freely Rotating Chain model, because the average of angle between $C_\alpha - C_\alpha$ bond has a relatively large standard deviation and it is not completely rigid (table II).

III. RESULTS

We performed equilibrium Monte Carlo simulation and we used the standard metropolis algorithm [22] with Pivot move [23] and a simple Hamiltonian of the form $H = \vec{f} \cdot \vec{r}$, where \vec{f} is the pulling force and \vec{r} is the position of the end of the chain while the other end is fixed at the origin. In the experiments, proteins are pulled slowly so we could use equilibrium Monte Carlo. We used the original random number generator of GNU C/C++ compiler to perform the simulation.

We started from a force of zero and an extended structure. We performed the simulation until reaching equilibrium, then we increased the force gradually. At each step we let the system reach the equilibrium, then we measured the average extension along the direction of the force. This process is equivalent to the AFM experiments with a soft cantilever [24],

| Atom-Atom | Distances (nm) | σ |
|----------------------------|----------------|----------|
| $C_{\alpha 1}C_{\alpha 2}$ | 3.82 | 0.05 |
| $C_{\alpha 1}N_1$ | 2.43 | 0.05 |
| $C_{\alpha 1}O_1$ | 2.40 | 0.02 |
| $C_1C_{\alpha 2}$ | 1.53 | 0.01 |
| $C_{\alpha 1}C_1$ | 1.53 | 0.01 |
| C_1N_1 | 1.33 | 0.04 |
| $N_1C_{\alpha 2}$ | 1.47 | 0.01 |
| N_1C_1 | 2.47 | 0.04 |
| C_1C_2 | 3.2 | 0.2 |

TABLE I: Atomic distances of the main atoms in peptides. Indexes show the order of atoms from left to right in the Fig. 1.

| $\vec{A} \cdot \vec{B}$ | Dot products | σ | Angle (Degree) |
|---|--------------|----------|----------------|
| $C_{\alpha 1}\vec{C}_{\alpha 2}.C_{\alpha 2}\vec{C}_{\alpha 3}$ | 0.19 | 0.25 | 78.9 |
| $N_1\vec{C}_{\alpha 2}.C_2\vec{C}_{\alpha 2}$ | 0.349 | 0.032 | 69.60 |

TABLE II: Angles in peptide units. The vectors are normalized and the indexes show the order of atoms from left to right in Fig. 1. You can see that FRC is not valid because the Standard Deviation of $C_{\alpha 1}\vec{C}_{\alpha 2}.C_{\alpha 2}\vec{C}_{\alpha 3}$ is too large compare to its average.

so we can use the experimental data of this Ref. [1].

We continued the process until we got a fully extended chain (~ 32 nm for 89 residues). We repeated it for chains with different numbers of residues. We found that the extension is proportional to the number of residues, independent of the force. In Fig. 3 one can see that all the extension-force curves collapse very well when their extensions have been divided by the number of residues (N). Because of this universality, it is useful to find a suitable fitting function to represent results of simulation and reproduce them for later uses. The Logistic Dose-Response,

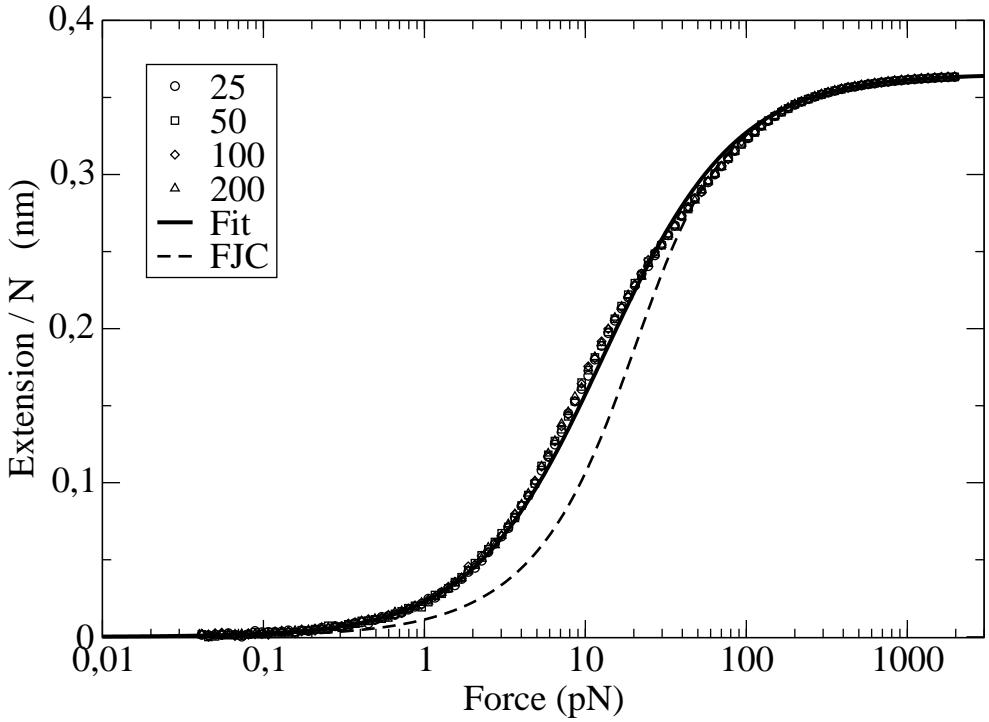


FIG. 3: Results of simulations for $N = 25, 50, 100$ and 200 peptides. All the curves have been divided by N . The curves collapse perfectly. The solid line shows Eq. 2 and the related parameters are in table III.

$$x = \frac{aN}{1 + (f_c/f)^\gamma}, \quad (2)$$

is a good candidate. Although it is simple and has only three parameters, it includes all the important features of the curves: a parameter to control the transition height (aN), another one to control the transition center (f_c) and one more to control the transition sharpness (γ). In Eq. 2 variables x , f and N are extension, force and the number of residues, respectively. The share of each peptide units in total length is $a = 0.3640$ nm and $f_c = 12.46$ pN shows the critical force which separates the swollen and the extended phases. The last parameter is the fitting exponent $\gamma = 1.021$ that controls the transition sharpness. If we want to fit the function to an experiment, we have to consider also the possible offset from zero.

The function 2 holds for the stretching of a chain without self interaction, but we can use it to fit the saw-tooth pattern as well. To do this, we consider that the multi domain chain can have both the unfolded entropic chains and the folded domains. We know the behavior

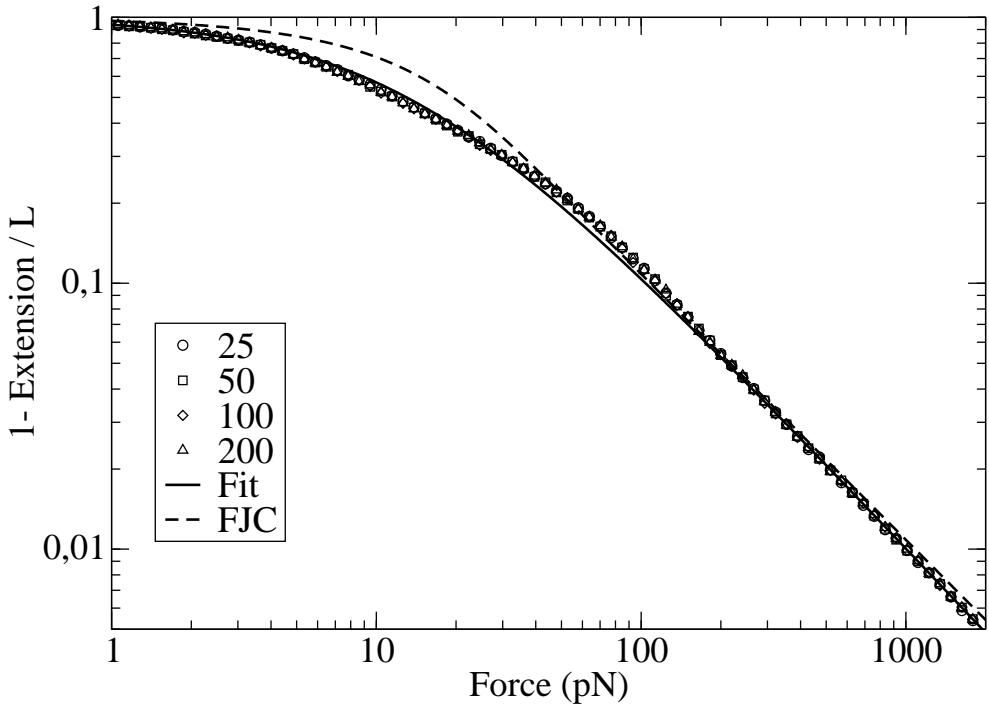


FIG. 4: Reduced extension versus force.

| | |
|----------|------------------------|
| a | 0.3651 ± 0.0004 nm |
| f_c | 13.0 ± 0.1 pN |
| γ | 1.072 ± 0.004 |
| R_d | 5.4 ± 0.3 nm |

TABLE III: Parameters of Eq. 3. Where a , f_c and γ has been calculated by simulation, R_d has been calculated by fitting the simulation results to that of saw-tooth like patterns.

of the former from Eq. 2, and we can approximate the latter as a rigid and inextensible objects of an unknown size. We will find its size from the peak to peak distance of the saw-tooth patterns. In fact, the distance between two consecutive peaks in saw-tooth patterns is equal to the length of an unfolded domain minus its end-to-end distance when it is folded. The following

$$x = \frac{anN}{1 + (f_c/f)^\gamma} + (n_{Ig} - n)R_d + x_0 \quad (3)$$

gives us the extension after the n th unfolding event. Here N is the number of residues

in a single domain, and n_{Ig} is the number of Ig repeats in the experiment. The parameter R_d shows the distance between the first residues of two consecutive folded domains which is equal to the end to end distance of a folded Ig domain plus the length of one peptide unit. The last parameter x_0 is the offset from zero in experiments.

The first term in Eq. 3 is the extension of unfolded domains and the second term is the contribution of folded domains to the contour length. It should be mentioned that we have assumed that the folded domains are rigid objects which only increase the length of the chain and do not contribute to the entropy and the force; therefore, we expect that the simulations fit better to the last events, when most of the domains are unfolded.

We know N and we have found a and f_c by simulation; therefore, only R_d remains unknown and can be used as a fitting parameter to set the peak to peak distance. We will have a good fit if we choose R_d equal to 5.4 nm; therefore, it gives us the end to end distance of a single folded domain that is approximately 5.0 nm. This value roughly agrees with the 4.3 size taken by NMR[7]. This small discrepancy (0.7 nm, almost twice of a peptide unit) might be due to the structure deformation of domains under tension.

Figure 5 displays the results of simulations and their comparisons to experimental. The experimental data has been taken directly from the graph of Ref. [1]. The parameters are displayed in table III.

IV. CONCLUSION

We used a model that has only one free parameter: the number of residues. By using this model, we could fit the result of the simulation to a single unfolding event accurately. This shows that the primary structure of the protein is not important in the entropic regime, within the experimental precision.

Our results suggest that this model is accurate enough to study the protein backbones in the entropic regime. Simpler models will need to introduce fitting parameters to the model to reproduce the experimental data.

Since our model performed well in describing the mechanical properties of protein backbones in the entropic regime, one may go one step further and use it as a base for studies like [25] in which our model can reduce the number of effective interactions and present a clearer and more straightforward picture. Besides, it is easy to consider also the hydrogen

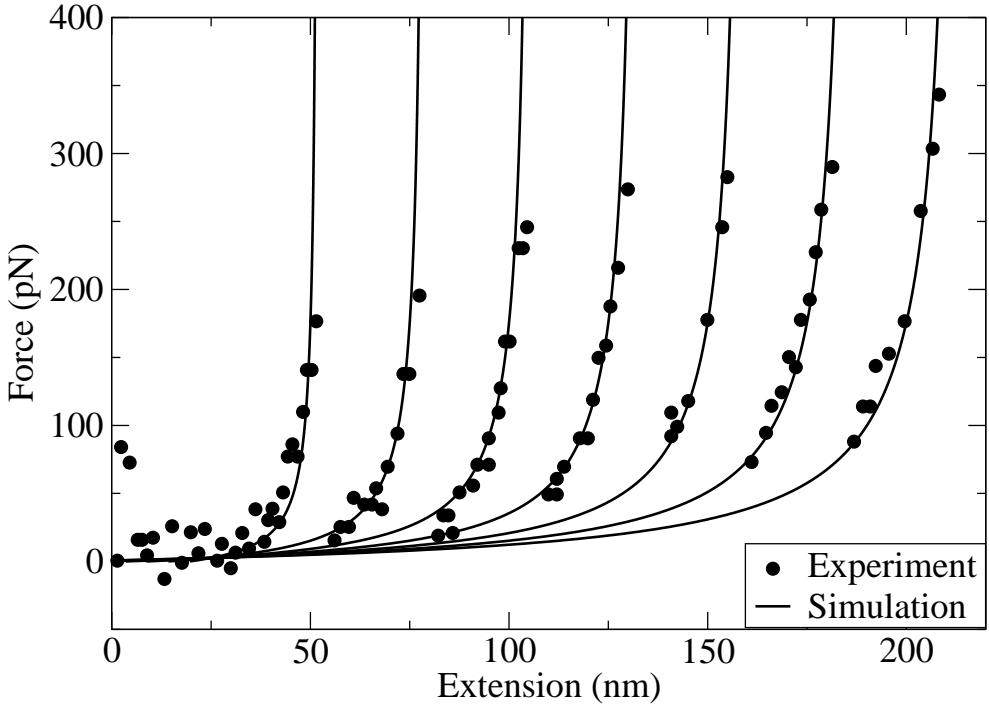


FIG. 5: We fitted the simulation (lines) to the experiment (circles) of [1] by following the concepts of the Eq. 3. Related fitting parameters are in table III.

bonds along the backbone, since it is sufficient to include oxygen and hydrogen atoms of the backbone.

[1] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, *Science* **276**, 1109 (1997).

[2] M. S. Z., Kellermayer, S. B. Smith, H. L. Granzier, and C. Bustamante, *Science* **276**, 1112 (1997).

[3] L. Tskhovrebova, J. Trinick, J. A. Sleep, and R. M. Simmons, *Nature* **387**, 308 (1997).

[4] R. B. Best, B. Li, A. Steward, V. Daggett, and J. Clarke, *Biophys Journal* **81**, 2344 (2001).

[5] T. E. Fisher, P. E. Marszalek, and J. M. Fernandez, *Nature structural biology* **7**, 719 (2000).

[6] S. Labeit and B. Kolmerer, *Science* **270**, 293 (1995).

[7] S. Imrota, A. Politou, and A. Pastore, *Structure* **4**, 323 (1996).

[8] E. Paci and M. Karplus, *Proc Natl Acad Sci USA* **97**, 6521 (2000).

- [9] M. Cieplak, T. X. Hoang, and M. O. Robbins, *Proteins* **56**, 285 (2004).
- [10] D. K. Klimov and D. Thirumalai, *Proc Natl Acad Sci USA* **96**, 6166 (1999).
- [11] N. M. Toan, D. Marenduzzo, and C. Micheletti, *Biophysical Journal* **89**, 80 (2005).
- [12] H. Kramers, *Journal of chemical physics* **14**, 415 (1946).
- [13] P. J. Flory, *Statistical Mechaics of Chain Molecules* (Wiley, New York, 1969).
- [14] O. Kratky and G. Porod, *Rec. Trav. Chim.* **68**, 1106 (1949).
- [15] J. Marko and E. Siggia, *Macromolecules* pp. 8759–8770 (1995).
- [16] A. Rosa, T. Hoang, D. Marenduzzo, and A. Maritan, *Macromolecules* **36**, 10095 (2003).
- [17] G. N. Ramachandran and V. Sasisekharan, *advanced in protein chemistry* **23**, 283 (1968).
- [18] L. Livadaru, R. R. Netz, and H. J. Kreuzer, *Macromolecules* **36**, 3732 (2003).
- [19] S. Takada, Z. Luthey-Schulten, and P. G. Wolynes, *J. Chem. Phys.* **110**, 11616 (1999).
- [20] A. V. Smith and C. Hall, *Proteins Structure Function and Genetics* **44**, 344 (2001).
- [21] F. Ding, J. Borreguero, S. Buldyrey, H. Stanley, and N. Dokholyan, *Proteins Structure Function and Genetics* **53**, 220 (2003).
- [22] N. Metropolis, *Journal ofChemical Physics* **21**, 1087 (1953).
- [23] N. Madras and A. Sokal, *Journal of Statistical Physics* **50**, 109 (1988).
- [24] H. J. Kreuzer, S. H. Payne, and L. Livadaru, *Biophysical Journal* **80**, 2505 (2001).
- [25] J. R. Banavar, T. X. Hoang, A. Maritan, F. Seno, and A. Trovato, *Physical Review E* **70** (2004).